·1·

Introduction

D. Crisan and B. Rozovskiĭ

The increasing complexity of the modern world makes the analysis and synthesis of high-volume data an essential feature in many real problems. Such analysis is addressed by different disciplines and from different perspective and, among them, nonlinear filtering features prominently. Nonlinear filtering is distinguished from other approaches by its probabilistic (in particular, Bayesian) nature. It is a field that combines aspects of stochastic analysis, information theory, and statistical inference. To date, nonlinear filtering is a mature theory that continues to expand by leaps and bounds. The breadth of its applications, firmly established and still emerging, is simply astounding. Early applications of nonlinear filtering such as cryptography, tracking, and guidance were mostly of military nature. Since then, nonlinear filtering has became an extremely potent tool in speech recognition, image and video processing, genetics, financial modeling, Bayesian networks, etc.

The celebrated Kalman–Bucy filter, designed for linear dynamical systems with linearly structured measurements, is probably the most famous Bayesian filter. Its generalizations to nonlinear systems and/or nonlinear observations are collectively referred to as *nonlinear filtering* (NLF). To put it succinctly, nonlinear filtering is an extension of the Bayesian framework to the estimation, prediction, and interpolation of nonlinear stochastic dynamics. Its output is the distribution of the estimated process (the "signal") given the data (the "observations") available. This distribution is commonly known as the posterior distribution of the estimated process. It is a theoretically optimal algorithm in that it provides the best estimate for the quantity of interest, more precisely, it minimises the mean square error of the estimator.

An important special case of the NLF paradigm that addresses Markov type dynamics is often referred to as Hidden Markov Models (HMM). It is an elegant and illuminating example which illustrates the principles of nonlinear filtering so it is worthwhile explaining it here.

Consider the following model: Let $(X_t,\ \ t = 0, 1, 2, \ldots)$ and $(Y_t, t = 0, 1, 2, \ldots)$ be two random sequences called, respectively, *state* and *observations*. The state is not directly observable. It is modeled as a Markov chain with the transition probability kernel $Q_t\,(x, y)$ and the initial distribution π_0. The observation sequence is related to the state by the formula

$$Y_t = h(X_t) + \xi_t,$$

where ξ_t is random noise. The a priori information contained in this model consists of the prior distribution π_0, the transition probability kernel $Q_t(x, y)$, and the distribution of ξ. The term "hidden Markov model" alludes to the fact that the Markov chain X_t is hidden from the observer by a possibly nonlinear transformation $h(\cdot)$ and the noise ξ_t. The task of nonlinear filtering is to compute, at each time t, the *posterior* distribution $\pi_{t|t}$ of the state X_t, in other words, the conditional distribution of X_t given the observation $Y_{0|t} = (Y_0, \ldots, Y_t)$. In this setting, the posterior distribution satisfies a two-step recursion:

$$\begin{aligned} &\text{prediction}: \quad \varphi_{t|t-1}(x) = \int Q_t(x', x)\, \varphi_{t-1|t-1}(dx') \\ &\qquad\qquad\qquad \text{and} \\ &\text{correction}: \pi_{t|t}(x) = \Psi_t(x)\, \pi_{t|t-1}(x) \Big/ \int \Psi_t(x')\, \varphi_{t|t-1}(dx'), \end{aligned} \tag{1.1}$$

where $\Psi_t(x) = g_t(\mathbf{y}_t - H(x))$ is the likelihood function. The first step consists in computing the conditional distribution of the state X_t given all but the last observation, i.e., given $(Y_0, \ldots, Y_{t-1})$. The second step is the well-known Bayes rule.

The early history of HMMs is shrouded in secrecy due to potential applications in cryptography. There exists an anecdotal evidence that the work on this topic was done by engineers in the early sixties but did not really come out in the open until the HMMs were "rediscovered" in late sixties. The crucial part of the developed "technology" is usually referred to as Baum–Welch algorithm (see papers [2] and [21]), particularly for frameworks where the state can take can have only a finite number of values (finite state space).

In the continuous setting, the prediction and correction steps merge and the posterior density $\pi_t(x) = \pi_{t|t}(x)$ is given by the Bayes formula.

$$\pi_t(x) = \frac{\varphi_t(x)}{\displaystyle\int_t \varphi_t(x')\, dx'}. \tag{1.2}$$

The unnormalized posterior density $\varphi_t(x)$ is a solution of the following equation:

$$\varphi_t(x) = \pi_0(x) + \int_0^t A^*(s, x)\, \varphi_s(x)\, dt + \int_0^t h_s(x)\varphi_s(x)\, dY_s, \tag{1.3}$$

where A^* is the dual of the generator of the signal's transition probability kernel $Q_t(x, y)$ and Y is the observation process. In particular, if the signal is given by the noisy kinematic equation

$$\dot{x}_t = a(t, x_t) + \sigma \dot{w}_t,$$

where $\dot{w}_t$ is white noise, then A^* has the form

$$A\varphi\,(x) = \frac{\sigma^2}{2}\varphi_t''\,(x) - (a\,(t,x)\,\varphi_t\,(x))'\,. \tag{1.4}$$

In the mid-sixties, the continuous time setting nonlinear filtering problem captured the attention of three young mathematicians: Duncan, Mortensen, and Zakai. They studied aspects of nonlinear filtering as a natural generalization of the well known linear filtering results of Kalman and Bucy. In four separate papers (two of them being PhD dissertations), they derived stochastic differential equations for unnormalized posterior distribution of a state process modeled by a continuous time Markov process. In paper [22], the state process was modeled by a continuous time jump Markov process with countable state space. The other three papers ([6], [16], and [23]) dealt with diffusion type state processes. Equation (1.4) and its generalizations are often referred to as Duncan–Mortensen–Zakai equation.

On the other hand, the (normalized) posterior density $\pi_t\,(x)$ solves a nonlinear stochastic PDE, which is usually referred to as the Kushner equation (see [12] (a corrected version of [11])). A more general version of the Kushner equation was derived by A. Shiryaev ([19]). These papers triggered a surge of activities in NLF during the late sixties and early seventies. The results of this period in the development of NLF were summarized in the influential book by R. Liptser and A. N. Shiryaev [14].

Note that Duncan–Mortensen–Zakai and Kushner equations are *stochastic partial differential equations* (SPDEs). In the sixties and seventies, SPDEs constituted a completely new subject for the stochastic community which quickly became an active area of research. Thus, one unintended but very important effect, triggered by the introduction of Duncan–Mortensen–Zakai and Kushner equations, was the fast development of a general theory of SPDEs. The first comprehensive accounts of these developments were published in [17] and [18]. For more details see the contributions by Krylov and Kunita in this volume.

Another milestone in the evolution of the theoretical side of NLF was related to the introduction of martingale techniques in the paper [7] by Fujisaki, Kallianpur and Kunita (more details can be found in Kunita's contribution in Part I of the Handbook). The martingale approach made it possible to deal with very general and diverse models of state and observations. Later on, Grigelionis and Mikulevicius extended the Fujisaki–Kallianpur–Kunita theory to even larger set of processes, including processes with irregular trajectories (see their contribution in Part I of the handbook).

In the last twenty years, a very impressive progress was made also in the study of asymptotic properties of SPDEs and, in particular, stability of nonlinear filters. This field is covered in great detail in Part III of the handbook.

From the very beginning, Bayesian filtering, both linear and nonlinear, has been an applied field. Numerous practical applications of Kalman filter and Baum–Welch algorithm are well documented.

The simplicity of the nonlinear filtering algorithm, particularly in the discrete setting, is deceptive. While being extremely effective and stable, it is computationally expensive. Both the prediction and the correction step involve computing integrals over the state space and these computations have to be executed every time new observations arrive. Moreover, in many important applications the computations have to be done in real time. Clearly, this is a serious complication since direct quadrature methods are effective in real time only when the dimension D of the state process is comparatively low (in general D should be no larger than 3). In contrast, the Kalman filter can deal in real time with hundreds of states. However, ad hoc extensions of linear Gaussian filters to the nonlinear setting such as the Extended Kalman Filter, are usually unsuccessful.

Fortunately, by the end of eighties two important factors have emerged: (a) a massive increase of computing power; and (b) the proliferation of Bayesian methodology into Monte Carlo simulations and vice versa.

For example, when implementing the Baum–Welch algorithm, one replaces the quadratures in the two steps of (1.1) by two Monte Carlo procedures. Nonlinear filtering algorithms based on Monte Carlo averaging are often called sequential Monte Carlo methods (SMCM) or particle filters. Such algorithms approximate of the posterior distribution using the empirical distribution of a system of n particles which evolve (mutate) according to the law of of the state. After each mutation the system is corrected: each particle is replaced by a random number of particles whose mean is proportional to the likelihood of the position of the particle. The most popular method for correcting the system is to sample with replacement n times from the empirical distribution of the population of particles weighted by their normalized likelihoods. The following is a simple algorithmic description of a garden variety sequential Monte Carlo method:

1. **Initialization** $[t = 0]$.
   For $i = 1, \ldots, n$, sample $x_0^{(i)}$ from π_0.
2. **Iteration** $[t - 1$ to $t]$.
   Let $x_{t-1}^{(i)}$, $i = 1, \ldots, n$ be the positions of the particles at time $t - 1$.
   (a) For $i = 1, \ldots, n$, sample $\bar{x}_t^{(i)}$ from $Q_{t-1}(x_{t-1}^{(i)}, x)dx$. Compute the (normalized) weight $w_t^{(i)} = g_t(\bar{x}_t^{(i)})/(\sum_{j=1}^n g_t(\bar{x}_t^{(j)}))$.
   (b) Pick $x_t^{(i)}$ by sampling with replacement from the set of particle positions $(\bar{x}_t^{(1)}, \bar{x}_t^{(2)}, \ldots, \bar{x}_t^{(n)})$ according to the probability vector of normalized weights $(w_t^{(1)}, w_t^{(2)}, \ldots, w_t^{(n)})$, $i = 1, \ldots, n$.

The approximation π_t^n of the posterior distribution is $\pi_t^n = \frac{1}{n}\sum_{i=1}^{n}\delta_{x_t^{(i)}}$, where $x_t^{(i)}$ for $i = 1, \ldots, n$ are the positions of the particles obtained after the second step of the iteration.

Following part (a) of the iteration, each particle changes its position according to the transition kernel of the signal. Step (a) of the algorithm is known as the *importance sampling* step (popular in the statistics literature) or *mutation* step (inherited from the genetic algorithms literature).

Step (b) of the iteration is called the *selection* step. The particles obtained after the first step of the recursion are multiplied or discarded according to the magnitude of the likelihood weights. In turn, the likelihood weights are proportional to the likelihood of the new observation given the corresponding position of the particle. The net effect of part (b) of the iteration is that it discards particles in unlikely positions and multiplies those in more likely ones. The particle filter with this choice of offspring distribution is called the *bootstrap filter* or the *sampling importance resampling* algorithm (SIR algorithm). It was introduced by Gordon, Salmond, and Smith in [8]. Within the context of the bootstrap filter, the second step is called the *resampling* step.

The theory of nonlinear filtering was well prepared for assimilating and expanding the particle approach. In particular, the Lagrangian approach to NLF, developed in the late 1970s and early 1980s (see [13], [10]), turned out to be a helpful framework for particle filters. The Lagrangian representations for forward and backward dynamics of the optimal filter, often called averaging over characteristics formulae, generalize the famous Feynman–Kac representation of solutions for deterministic parabolic equations. The averaging in these formulas is conditioned on the available observations. The Lagrangian characteristics model the stochastic dynamics of the Monte Carlo particles.

Various versions of the optimal nonlinear filters based on Monte Carlo resampling were developed in the 1990s, in particular, interacting particle filter, sampling/importance resampling particles filter, branching particles filter, etc. For a review, see [5], [4], [1] and the contributions in Part VII of the handbook.

The introduction of particle filters has influenced fundamentally the area of nonlinear filtering. It has extended the reach of NLF to higher dimensional applications and, therefore, enlarged the range of practical applicability of nonlinear filtering. It has also posed many new problems and opened new avenues of research. In our opinion, it is the most important development in the last two decades of research in nonlinear filtering.

We complete the introduction with a description of the contributions comprising the handbook.

The first two parts of the handbook contain classical theoretical results related to the filtering equations covered by six contributions. They are as follows:

The contribution of Kunita is a two-part introduction to nonlinear filtering. In the first part, the filtering problem in discrete time is analysed together with some preliminary results on stochastic calculus required for the study of continuous time nonlinear filtering. The Bayes formula for the computation of the nonlinear filter is derived and the existence of the innovation process is discussed. In the second part, the filtering equations are derived for a general class of system processes having the semimartingale property and the Cauchy problem associated with these equations is studied: existence, uniqueness, and smoothness results are included.

The filtering problem corresponding to processes with jumps is analyzed in the chapter of Grigelionis and Mikulevicius. The existence and uniqueness of the solutions of the filtering equations are discussed. Some examples and an application in financial mathematics (volatility tracking) are included.

In the contribution of Kurtz and Nappo, the signal process is defined as the solution of a martingale problem and its conditional distribution with respect to the observation filtration is recast as the solution of a *filtered* martingale problem. It is shown that uniqueness for the signal's martingale problem implies uniqueness for the filtered martingale problem which in turn implies the Markov property for the conditional distribution considered as a probability measure-valued process. Other applications include a Markov mapping theorem and uniqueness for the filtering equations.

In the chapter by Krylov, the smoothness in L_p sense of filtering densities is discussed. The filtering equations are normally considered in terms of formal adjoint of operators in nondivergence form. Here they are rewritten in the divergence form and the smoothness of solutions is established under very general conditions (Lipschitz continuity of the coefficients of the system).

Two different applications of the Malliavin calculus to nonlinear filtering are discussed in the contribution of Chaleyat-Maurel. The first one deals with the existence and smoothness of a density for conditional laws in filtering theory, whereas the second one is concerned with the problem of the existence or nonexistence of finite-dimensional filters. The two applications are different in nature. In the first one, the observation is considered as fixed, and the Malliavin calculus is applied to the signal in a finite-dimensional approach. In the second one, the Malliavin calculus is applied to the observation through the Zakai equation, which is a stochastic partial differential equation, and the setting is thus infinite dimensional.

The chapter by Lototsky discusses various methods of solving the nonlinear filtering problems using expansions of the optimal filter in the chaos space of the observation process. The elements of the expansion can be either multiple integrals or the Cameron–Martin basis. Two particular filtering algorithms are discussed for the time-homogeneous diffusion filtering model with possible correlation between the state process and the observation noise. Both algorithms rely on the Cameron–Martin version of the chaos expansion, and the approximate filter is a finite linear combination of the chaos elements generated by the observation process. The coefficients in the expansion depend only on the deterministic dynamics of the state and observation processes.

Part III of the handbook covers stability properties and asymptotic analysis of the filtering solution. It includes the following:

The chapter by Kleptsyna and Veretennikov presents some of their recent results for the filtering problem for which the signal has an unknown/unspecified initial condition. The authors show that, under suitable conditions, the filtering algorithm forgets its wrong initial data in the long run, that is, the difference between the conditional measures provided by the filtering algorithm with the exact and wrong initial data converges to zero in some suitable topology. Both the discrete and the continuous frameworks are discussed.

The contribution by Atar presents a review of tools from multiplicative ergodic theory and the theory of positive operators and their usage in the analysis of exponential stability of the optimal nonlinear filter. Particularly, in the case of finite state, the chapter studies the filter sensitivity to perturbations in its initial data and its relation to the Lyapunov spectral gap associated with the filtering equation. In a general setting, it is shown how the Hilbert's metric and Birkhoff's contraction coefficient are used to estimate the decay rate of the error.

The chapter by Chigansky, Liptser, and Van Handel presents a survey of some *intrinsic* methods for studying the stability of the nonlinear filter. These intrinsic methods are methods which directly exploit the fundamental representation of the filter as a conditional expectation through classical probabilistic techniques such as change of measure, martingale convergence, coupling, etc. These methods allow one to establish stability of the filter under weaker conditions compared to other methods, e.g., to go beyond strong mixing signals, to reveal connections between filter stability and classical notions of observability, and to discover links to martingale convergence and information theory.

The contribution of Budhiraja describes conditions under which the solution of the filtering problem satisfies the Feller property and the existence of invariant measures. It also studies the ergodicity of the nonlinear filter and gives some sufficient conditions, under which this property holds, phrased in terms of certain stability properties of the nonlinear filter, for example, the finite memory property or asymptotic stability.

The chapter by Stannat gives an overview on results of stability of the optimal filter for nonergodic signal processes with state space $\mathbb{R}^d$ observed with independent additive noise, both in discrete and continuous time. Explicit lower bounds on the rate of stability in terms of the coefficients of the signal and the observation are obtained, using a parabolic ground state transform with respect to log-concave measures of the recursive algorithm for the optimal filter. The lower bounds in the time-continuous case are obtained as limits of lower bounds for appropriate time-discrete approximations. As particular examples, the Kalman and the Kalman–Bucy filters and filters with signals induced by gradient-type stochastic differential equations are discussed.

Part IV of the handbook includes several special topics, which we describe briefly below:

The *pathwise* theory of filtering is discussed in the contribution of Davis. This theory is concerned with casting the filtering equations in a form in which the filtering estimates can be computed separately for each sample path of the observation process. The chapter presents a pathwise theory for the case where the signal is a diffusion on a finite-dimensional manifold and there is correlation with the observation noise. A geometric setting is natural for this problem, which also brings in Kunita's decomposition theorem for solutions of stochastic differential equations and a family of observation-dependent multiplicative functionals of the signal process.

It turns out that the observation process can be decomposed into two components. One is the integral of the expectation of a function of the signal conditioned with respect to the observation data. The second one is a Brownian motion adapted to the observation filtration, called the innovation process. The natural filtration of the innovation process is included in, but not necessarily equal to the observation filtration. Establishing the cases where the two filtrations are equal has become known as the innovation problem. The contribution of Heunis establishes conditions under which the two filtrations are equal.

The paper of Duncan discusses some results for nonlinear filtering problems where the processes satisfy stochastic differential equations driven by fractional Brownian motions. Fractional Brownian motions are a family of Gaussian processes that include the standard Brownian motion and that seem to be appropriate models for many physical phenomena. The paper covers properties of the family of fractional Brownian motions and the explicit expressions for the Radon–Nikodym derivatives appearing in the formulae for the solution of the filtering problem. It also contains: a stochastic integral equation for the evolution of the conditional expectation of a function of the state process, results for the prediction of processes generated by a fractional Brownian motion and two relations between filtering and mutual information.

Part V of the handbook covers the topics of estimation and control in nonlinear filtering.

The contribution of Newton investigates nonlinear filtering from an information theoretic viewpoint. At its heart are two distinct dualities: one is a feature of time reversal, the other is an instance of an abstract Fenchel–Legendre transform for Bayesian estimators. The first duality arises from a time symmetry in the joint dynamics of the signal process and its nonlinear filter. The second duality is that between the *full information* of a log-likelihood function and the *information gain* of the corresponding posterior distribution in the context of Bayesian estimation. By applying this duality to the path estimation problems associated with the forward and reverse time filters, forward and backward stochastic optimal control problems are obtained, in which the two filters appear in the value functions. The second duality, applied in this way, becomes the duality between estimation and control.

The contribution of Bensoussan, Cakanyıldırım, and Sethi develops a general filtering framework for the problem of estimating the state of a system whose dynamics are governed by a discrete-time Markov process. The chapter presents a number of applications to inventory control systems with partial observations. The authors show how one can transform the nonlinear transition equations into linear ones. This transformation facilitates considerably the study of the associated control problem and the corresponding Bellman equation in a convenient functional space.

The contribution of Bar-Shalom and Blom studies stochastic hybrid systems. These are two component Markov processes $\{x_t, \theta_t\}$, where $\{\theta_t\}$ is a Markov chain and $\{x_t\}$ is the solution of a stochastic difference equation (SDE) whose coefficients depend of $\{\theta_t\}$. The chapter covers the exact Bayesian filter recursions and particle filter approximations for these two component Markov processes. During the development of the exact and particle filter recursions, a key role is played by the exact equations that form the basis of the Interacting Multiple Model filter.

Part VI of the handbook includes several topics related to the approximation theory for the filtering problem. The following are covered:

The filtering problem consists, in particular, in finding the best (in the sense of mean-square error) $\mathcal{Y}_t$-measurable estimator $\hat{X}_t$ of the signal X_t, that is the minimizer of the *filtering error*

$$P_t = \mathbb{E}\left[(X_t - \hat{X}_t)(X_t - \hat{X}_t)^T\right],$$

where $(X_t - \hat{X}_t)^T$ is the row vector associated to $(X_t - \hat{X}_t)$. Of course $\hat{X}_t = \mathbb{E}[X_t|\mathcal{Y}_t]$. Explicit expressions for $\hat{X}_t$ and, respectively, P_t are typically hard to obtain. The chapter of Zeitouni investigates a-priori bounds (both upper and lower) of the matrix P_t. Of particular interest are the bounds that are tight when the observation noise is small.

As stated above, the solution of the continuous time filtering problem can be represented as a ratio of two expectations of certain functionals of the signal process. These functionals are parametrized by the observation path $\{Y_s, s \geq 0\}$. However, in practical applications, only the values of the observation corresponding to a discrete time partition are available, i.e., $\left\{Y_{t_i}, i = 0, 1, \ldots\right\}$. This leads to an approximation of the filtering solution in terms of functionals parametrized by these discrete observations. The convergence rate of this approximation as a function of the partition mesh is studied in the contribution of Crisan. It is shown that the two critical factors that influence the order of convergence are the smoothness of the semigroup associated to the signal and the smoothness of the sensor function h.

The chapter by Le Gland, Monbet, and Tran discusses the ensemble Kalman filter (EnKF). Interpreting the ensemble elements as a population of particles with mean-field interactions, the authors prove the convergence of the EnKF, with the classical rate $1/\sqrt{N}$, as the number N of ensemble elements increases to infinity. In the linear case, the limit of the empirical distribution of the ensemble elements is the usual (Gaussian distribution associated with the) Kalman filter, as expected, but in the more general case of a nonlinear state equation with linear observations, this limit differs from the usual Bayesian filter.

Part VII covers the particle approach for solving the filtering problem. It includes the following contributions:

The contribution of Xiong is a survey of recent results on the particle system approximations to filtering problems in continuous time. Firstly, a weighted particle system representation of the optimal filter is given and a numerical scheme based on this representation is presented together with the convergence result to the optimal filter. Secondly, to reduce the estimation error due to the exponential growth of the variance for individual weights, a branching weighted particle system is defined and an approximate filter based on this particle system is included. Its approximate optimality is proved and the rate of convergence is characterized by a central limit type theorem. Thirdly, as an alternative approach in reducing the estimate error, an interacting particle system (with neither branching nor weights) to direct the particles toward more likely regions is proposed and the corresponding convergence result for this system is established. Finally, the weighted branching particle systems is used to approximate the optimal filter for the model with point process observations.

The contribution of Doucet and Johansen is a survey of results on particle system approximations for filtering problems in discrete time. Just as in the continuous time framework, optimal estimation problems for discrete nonlinear non-Gaussian state-space models do not typically admit analytic solutions. Since their introduction in 1993, particle filtering methods have become a very popular class of algorithms to solve these estimation problems numerically in an online manner, i.e. recursively, as observations become available. Particle filtering methods are now routinely used in fields as diverse as computer vision, econometrics, robotics and navigation. The objective of the contribution is to provide a complete, up-to-date survey of this field. Basic and advanced particle methods for filtering, as well as smoothing, are presented.

The contribution by Del Moral, Patras, and Rubenthaler presents a mean field particle theory for the numerical approximation of Feynman–Kac path integrals in the context of nonlinear filtering. The authors show that the conditional distribution of the signal paths given a series of noisy and partial observation data is approximated by the occupation measure of a genealogical tree model associated with mean field interacting particle model. The complete historical model converges to the McKean distribution of the paths of a nonlinear Markov chain dictated by the mean field interpretation model. The chapter also contains a review of the stability properties and the asymptotic analysis of these interacting processes, including fluctuation theorems and large deviation principles and a Laurent type and algebraic tree-based integral representations of particle block distributions.

The chapter by Schön, Gustafsson, and Karlsson contains a number of real-time applications of the particle filter (PF) in both the signal processing and the robotics communities. The authors present several applications to positioning of moving platforms detailing the experiences of using the PF in practice. The applications concern positioning of underwater vessels, surface ships, cars, and aircraft using geographical information systems containing a database with features of the surrounding. In the robotics community, the PF has been developed into one of the main algorithms (FastSLAM) for solving the simultaneous localization and mapping (SLAM) problem. This can be seen as an extension to the aforementioned applications, where the features in the geographical information system are dynamically detected and updated on the fly.

A key problem in filtering, which is only partially addressed by particle filters, is to maintain a good description of the evolving posterior measure using *minimal* computational effort. Recently, it has been shown that a new class of methods developed for approximation distributions of solutions of stochastic differential equations, collectively known as cubatures on Wiener space, can be used to approximate the conditional distribution in the filtering problem. The chapter by Litterrer and Lyons is a survey on cubature on Wiener space and some related algorithms. It also describes how recombination can be added to the basic algorithm as a way to control the number of particles in the approximation when the method is iterated.

Part VIII contains contributions related to numerical methods in nonlinear filtering. The contribution of Kushner considers two types of numerical algorithms for nonlinear filters. The first is based on the Markov chain approximation method, a powerful approach to numerical problems in stochastic control. It yields an approximation to the conditional density and converges in the weak sense as the approximation parameter goes to zero. Various forms are developed and both convergence and robustness results are included. The second type of approximation is called the assumed density approach, where one supposes that the conditional density takes a given parametrized form, and the evolution equations for the parameters are developed. Most typically, this assumed density is Gaussian (more rarely, a Gaussian mixture) and the parameters are the conditional mean and covariance. The method is heuristic, but has been shown to give good results for many problems.

The chapter by Hairer, Stuart, and Voss is an overview of the Bayesian approach to a wide range of signal processing problems in which the goal is to find the signal. In the case of ordinary differential equations (ODEs) this gives rise to a finite dimensional probability measure for the initial condition, which then determines the measure on the signal. In the case of stochastic differential equations (SDEs) the measure is infinite dimensional. The authors derive the posterior measure for these problems, applying the ideas to ODEs and SDEs, with discrete or continuous observations, and with coloured or white noise. The authors highlight the common structure inherent in all of the problems, namely that the posterior measure is absolutely continuous with respect to a Gaussian prior. This structure leads naturally to the study of Langevin equations which are invariant for the posterior measure and they highlight the theory and open questions relating to these S(P)DEs.

The contribution of Clark and Vinter is concerned with an important class of filtering problems referred to as tracking problems, where the objective is to estimate the state of a moving target from noisy sensor measurements. For many tracking problems of interest, the equations for the conditional distribution of the state are computationally intractable and the key challenges therefore relate to their approximation. The chapter identifies an important class of tracking problems, in which the nonlinearities involved in the models of the state and observations processes are confined to the observations process. Special cases involve bearings-only tracking, range-only tracking, and various tracking problems where measurements are suppressed or degraded in some "nonlinear" fashion. A general methodology is presented for constructing filters for such problems which typically provide superior estimates to those obtained by classical linearization techniques. The specific form taken by these filters in four cases of interest is examined in detail. In the case of bearings-only tracking, the filter is known as the shifted Rayleigh filter.

It is well known that numerical methods for nonlinear filtering problems, which directly use the Kallianpur–Striebel formula, can exhibit computational instabilities due to the presence of very large or very small exponents in both the numerator and denominator of the formula. The chapter by Milstein and Tretyakov introduces a class of computationally stable schemes by exploiting the innovation approach. The authors propose Monte Carlo algorithms based on the method of characteristics for linear parabolic stochastic partial differential equations. Convergence and some properties of the considered algorithms are studied and variance reduction techniques are discussed. The chapter also includes results of some numerical experiments.

The last part of the handbook includes a number of applications of nonlinear filtering in financial mathematics:

The chapter by Frey and Runggaldier considers filtering problems that arise in Markovian factor models for the term structure of interest rates and for credit risk. The connections with the filtering problem is based on the fact that investors act on the basis of only incomplete information about the factors. The current state of the factors has to be inferred/filtered from observable financial quantities. The main goal of the chapter is the pricing of derivative instruments in the interest rate and credit risk contexts, but also other applications are discussed.

The contribution of Elliott, Miao, and Wu introduces a generalized stochastic volatility model to help price energy-related assets by capturing two critical features: mean-reverting prices and a volatility which follows different dynamics in different states of the world. Assuming the dynamics of the states are represented by a hidden Markov chain, the authors apply filtering techniques and the EM algorithm to a time-series model for parameter estimation. Several new filters and closed form estimates for all parameters are derived in the paper. Applications of the proposed model in other fields of finance are also discussed.

The contribution of Pham is a survey of the methods involved in portfolio selection with partial observation. The author describes both the theoretical and numerical aspects related to these optimization problems. The presentation is divided in two parts. The first part covers the continuous-time problem: here, the mean rates of return of the asset prices are not directly observable. Investors observe only asset prices. By the method of change of probability and innovation process in filtering theory, the partial observation portfolio selection problem is transformed into a full observation one with the additional filter state variable, for which one may apply the martingale or PDE approach. The following cases for the modeling of the unobservable mean rate of return are investigated: Bayesian, linear-Gaussian, and finite-state Markov chain. The second part covers discrete-time optimization problems: this context includes the case of unobservable volatility. The numerical approximation of the optimization problem under partial observation is studied. Several numerical experiments illustrate the results for hedging problems in the context of partially observed stochastic volatility models.

The chapter by Scott and Zeng surveys the recent developments in a general filtering model with counting process observations for the micromovement of asset price and its related statistical analysis. The normalized and unnormalized filtering equations as well as the system of evolution equations for

Bayes factors are reviewed. A Markov chain approximation method is used to construct recursive algorithms and their consistency is proven. The authors employ a specific micromovement model built upon the model linear stochastic differential equation to show the steps to develop a micromovement model with specific types of trading noises. The model is further utilized to show the steps to construct consistent recursive algorithms for computing the trade-by-trade Bayes estimates and the Bayes factors for model selection.

References

[1] A. Bain and D. Crisan, Fundamentals of Stochastic Filtering. *Stochastic Modeling and Applied Probability* **60** (2009), Springer.

[2] L. E. Baum and T. Petrie, Statistical Inference for Probabilistic functions on finite state Markov chains. *Ann. Math. Stat.* **37** (1966), 1554–63.

[3] V. E. Benesh, Exact finite-dimensional filters for certain diffusions with nonlinear drift, *Stochastics* **5**, (1981), 65–92.

[4] P. Del Moral, *Feyman–Kac Formulae. Genealogical and Interacting Particle Systems with Applications*. Springer-Verlag, New York, (2004).

[5] A. Doucet, N. de Freitas, and N. Gordon (eds), *Sequential Monte Carlo Methods in Practice*, Springer-Verlag, New York, (2001).

[6] T. E. Duncan, Tech. Report 7001-4, Stanford University, Center for Systems Research, May (1967).

[7] M. Fujisaki, G. Kallianpur, and H. Kunita, Stochastic differential equations for the nonlinear filtering problem, *Osaka J. Math.* **9** (1972), no. 1, 19–40.

[8] N. J. Gordon, D. J. Salmon, and A. F. M. Smith. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEEE Proceedings*, Part F, pp. 107–113, (1993).

[9] R. E. Kalman and R. S. Bucy, New results in linear filtering and prediction problems, *J. Basic Engineering, Trans. ASME* **83D** (1961), 95–108.

[10] H. Kunita, *Stochastic Flows and Stochastic Differential Equations*, Cambridge University Press, Cambridge, (1982).

[11] H. J. Kushner, On the differential equations satisfied by conditional probability densities of Markov processes, with applications, *J. Soc. Indust. Appl. Math. Ser. A Control* **2** (1964), 106–19.

[12] H. J. Kushner, Dynamical equations for optimal nonlinear filtering, *J. Differential Equations* **3** (1967), 179–90.

[13] N. V. Krylov and B. L. Rozovskiĭ, Characteristics of second-order degenerate parabolic Itō equations. *Trudy Sem. Petrovsk.* **8** (1982), 153–68.

[14] R. Sh. Liptser and A. N. Shiryayev, *Statistics of Random Processes, I, II*, Springer, New York, (2002).

[15] S. V. Lototsky, R. Mikulevicius, and B. L. Rozovskiĭ, Nonlinear filtering revisited: a spectral approach, *SIAM J. Contr. Optim.* **35** (1997), no. 2, 435–61.

[16] R. E. Mortensen, *Tech. Report ERL–66–1*, UC Berkeley, Electronics Research Laboratory, Aug. (1966).

[17] E. Pardoux, Filtrage non linéaire et équations aux dérivées partielles stochastiques associées. In: *Ecole d'Eté de Probabilités de Saint-Flour XIX*, (1989).

[18] B. L. Rozovskiĭ, *Stochastic Evolution Systems. Linear Theory and Applications to Non-linear Filtering*, Kluwer Academic Publishers, Dordrecht, 1990 (trans. from the 1982 Russian original).

[19] A. N. Shiryaev, On stochastic equations in the theory of conditional Markov process, *Teor. Verojatnost. i Primenen.* **11** (1966), 200–6 in Russian; English trans. *Theor. Probability Appl.* **11** (1966), 179–84.

[20] R. I. Stratonovich, Conditional Markov processes, *Theor. Probab. Appl.* **5** (1960), 156–78.

[21] L. R. Welch, Hidden Markov Models and Baum–Welch Algorithm (The Shannon Lecture). *IEEE Information Theory Society Newsletter* **53**, no. 4, Dec. (2003)

[22] M. Zakai, *The optimal filtering of Markov jump processes in additive white noise*. Research Note No. 563. Applied Research Laboratory, Sylvania Electronic Systems, Waltham, Massachusets. (1965).

[23] M. Zakai, On the optimal filtering of diffusion processes, *Z. Wahrsch. Verw. Gebiete* **11** (1969), no. 3, 230–43.